International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Deep learning for multimodal perception: Improving lidar and radar integration in self-driving cars

Rijul Rajput *

Dougherty Valley High School 10550 Albion Rd, San Ramon, CA 94582, USA.

International Journal of Science and Research Archive, 2025, 16(01), 1396-1408

Publication history: Received on 03 June 2025; revised on 12 July 2025; accepted on 15 July 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.16.1.2100

Abstract

Autonomous driving relies on a range of advanced technologies, including computer vision, radar, lidar, and deep learning. Each of these components brings distinct strengths and limitations, but when used together, they create powerful synergies. While many self-driving systems employ radar or lidar independently, this separation often leaves critical gaps in perception. Combining radar and lidar with deep learning can significantly enhance a vehicle's environmental awareness by leveraging the complementary properties of each technology, thereby reducing individual weaknesses and improving overall safety. This paper explores the historical evolution and current challenges associated with deep learning, radar, and lidar in autonomous systems. It also examines methods for integrating radar and lidar data through a deep learning approach called multimodal learning, focusing on two key frameworks: M2-Fusion and ST-MVDNet. Although prior research on radar-lidar fusion remains limited due to the ongoing advancement of these technologies, various integration strategies have been introduced.

The objective of this paper is to present a systematic review of lidar, radar, and deep learning, offering a cohesive summary of their technological development. We also provide a critical evaluation of the M2-Fusion and ST-MVDNet models, highlighting their potential as well as the limitations that may affect their real-world implementation in modern autonomous vehicles.

Keywords: Autonomous Vehicles; Radar-Lidar Fusion; Deep Learning; Multimodal Learning; Object Detection; Fault Tolerance; M2-Fusion

1. Introduction

In autonomous driving, computer vision serves as the foundational pillar of the system's ability to perceive and interpret its environment. Much like how blindness impairs a human's ability to drive, the absence of computer vision prevents an autonomous vehicle from navigating effectively. These systems incorporate various components of computer vision, including deep learning, radar, and lidar, to detect and classify objects without human input.

Deep learning, a specialized branch of machine learning, involves training neural networks to identify complex patterns in data, enabling high-accuracy performance across a variety of tasks [1]. The field has advanced rapidly with the emergence of Large Language Models (LLMs) and generative tools for image and video creation. Radar (radio detection and ranging), while a much older technology, plays a key role in computer vision by using radio waves to detect objects and measure their distance, speed, and direction [2]. In contrast, lidar (light detection and ranging) is a newer technology that uses laser light to capture high-resolution spatial data, creating precise 3D maps of the environment [3].

* Corresponding author: Rijul Rajput

Radar excels in adverse weather conditions and long-range detection, making it a dependable component in navigation systems. Lidar, on the other hand, offers superior precision and the ability to generate detailed environmental maps, making it valuable for autonomous vehicles, robotics, and environmental monitoring. Despite growing research on integrating radar and lidar, a comprehensive study that analyzes the synergy between the two, assessing their complementary strengths and weaknesses, and reviews contemporary integration methods is still lacking. Our work addresses this gap through a synergistic analysis aimed at enhancing the use of radar and lidar for object detection in autonomous systems. By applying artificial intelligence, specifically deep learning, the interpretation of sensory data can be refined, enabling reliable and intelligent decision-making for autonomous driving. In this paper, we analyze two recent deep learning approaches, M2-Fusion and ST-MVDNet, that integrate radar and lidar data for improved object detection in self-driving vehicles.

2. Deep learning introduction

Deep learning is a subfield of machine learning that focuses on using multilayered neural networks to generate outputs based on complex input data. The roots of deep learning trace back to 1943, when Walter Pitts and Warren McCulloch published their seminal paper, *"A Logical Calculus of the Ideas Immanent in Nervous Activity"* [4]. In this foundational work, they likened the brain to a computing machine, with neurons functioning as individual processors. They introduced the first computational model of a neuron, which accepted an input ggg and produced an output f [Figure 1]. This model mirrors the structure of biological neurons, incorporating a dendrite to receive signals and an axon to transmit the output.

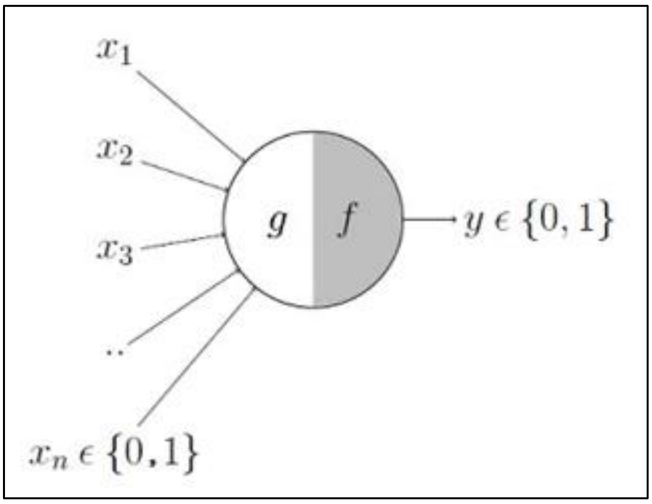

Figure 1 The first computational model of a neuron, as proposed by Pitts and McCulloch[5]

Deep learning leverages neural networks, which are powerful tools capable of solving a wide variety of complex problems. These networks consist of multiple layers of interconnected neurons that process data through weighted connections and biases [Figure 2]. In the context of autonomous driving, neural networks are used for critical functions such as street sign recognition, lane detection and tracking, and automated parking assistance. By automating many routine tasks traditionally handled by human drivers, deep learning contributes to a more efficient, comfortable, and safer driving experience [2].

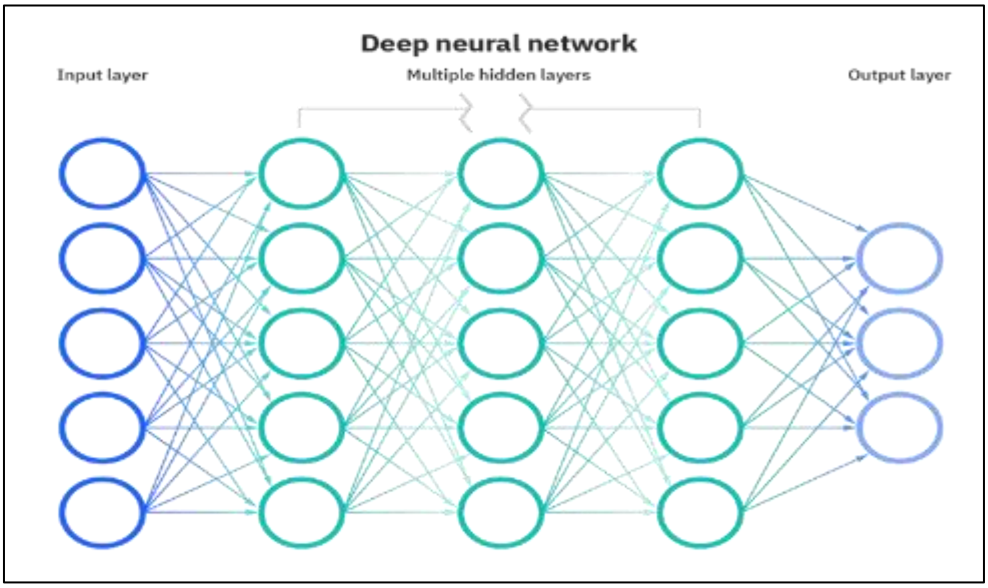

Figure 2 The structure of a deep neural network involves an input layer (left), several hidden layers that process information (middle), and an output layer (right)[6]

SAE International established a six-level classification system known as the SAE Levels of Automation to define the stages of autonomous driving [7]. The first three levels, starting from level 0, represent driver assistance technologies. For instance, level 0 includes features like automatic emergency braking, level 1 introduces lane centering, and level 2 combines adaptive cruise control with lane centering to provide partial driving automation. The next three levels (levels 3 through 5) shift control entirely to the vehicle, ranging from conditional automation to full autonomy. At level 4, vehicles can operate without a steering wheel or pedals in certain conditions, while level 5 envisions cars that can drive anywhere under any circumstance without human intervention. As these levels progress, the complexity and capability of the system increase significantly. Several features outlined in the SAE framework, such as "local driverless taxi" and "traffic jam chauffeur", are currently being tested. Industry leaders like Waymo and Tesla are at the forefront of implementing these advanced functionalities. Waymo has successfully developed a system capable of transporting passengers without a human driver, aligning with level 4 automation.

Meanwhile, Tesla's Full Self-Driving (FSD) technology uses neural networks to eliminate the need for human input, pushing the boundaries of autonomous navigation [8]. These advancements demonstrate how deep learning continues to drive significant progress in the development of self-driving vehicles. Despite their innovations, current autonomous systems have yet to achieve SAE level 5 autonomy, as they still face significant limitations. One of the primary methods for generating spatial awareness in autonomous vehicles involves the use of lidar and camera technologies [9]. However, these sensors are susceptible to performance degradation in adverse weather conditions such as fog, rain, or snow. To address this vulnerability, radar is often integrated as an additional modality, offering enhanced reliability and robustness.

The proposed deep learning-based approaches, M2-Fusion and ST-MVDNet, utilize radar inputs to augment object detection capabilities when lidar and camera data are compromised. These methods allow modern autonomous vehicles to edge closer to levels 4 and 5 by leveraging deep learning for feature extraction across multiple sensor types. This showcases the critical role deep learning plays in enhancing perception systems through multimodal fusion. In recent years, computer vision and deep learning have become closely intertwined, primarily due to the success of convolutional neural networks (CNNs) in image recognition and segmentation tasks. Originally introduced by Yann LeCun, CNNs are designed to process grid-like data structures and utilize stacked layers, such as convolutional layers with learnable filters and pooling layers to reduce dimensionality, to capture and analyze spatial features in image data. These architectures allow vehicles to detect and interpret complex objects, including partially obscured road signs and traffic cones [10]. In the M2-Fusion model, a CNN generates a feature map that aligns and links object features between radar and lidar 3D point clouds, improving detection accuracy [Figure 3].

Nevertheless, integrating deep learning into autonomous vehicles introduces challenges. First, building models that can generalize well across the unpredictable variety of real-world driving scenarios is inherently complex [11]. Second, the rise of deep learning has also brought increased vulnerability to adversarial attacks, where small, often imperceptible changes in input data can mislead models and alter decisions [11], [12]. To address these risks, developers are exploring strategies such as edge computing and privacy-preserving learning techniques. For example, federated learning allows models to be trained locally on edge devices without transmitting large volumes of sensitive data to central servers, thereby enhancing privacy and reducing attack surfaces. As deep learning continues to evolve and push the boundaries of software innovation in autonomous driving, complementary advancements in hardware, particularly in radar and lidar, are progressing in parallel, strengthening the ecosystem of safe and intelligent self-driving systems.

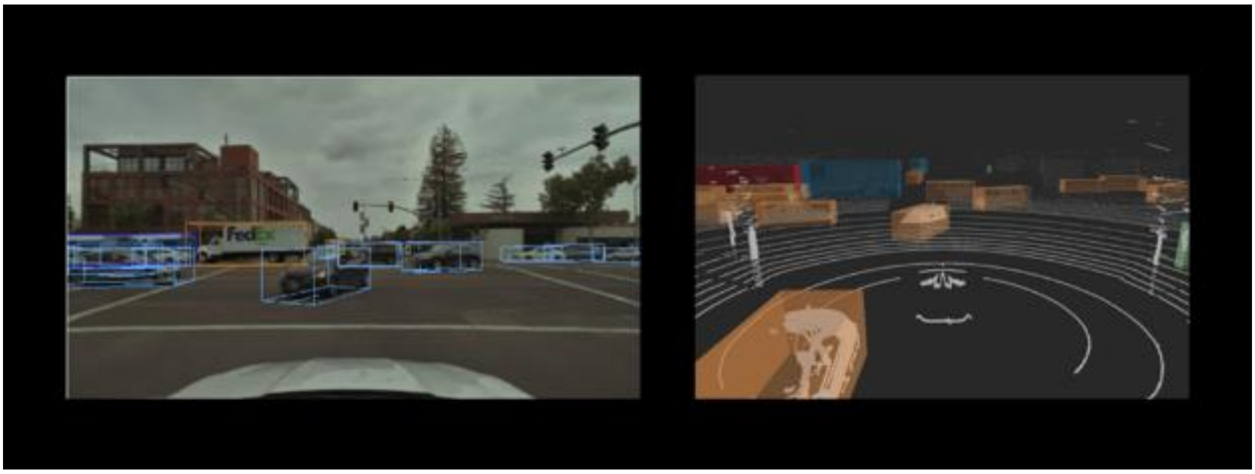

Figure 3 Lyft's lidar system detects different types of vehicles and categorizes them into separate classifications with CNNs; trucks and larger automobiles are shown differently in contrast to the orange cars in the right graphic[13]

3. Radar methods

Originally invented in 1935 to help pilots avoid thunderstorms, radar remains a widely respected and versatile technology still used in various applications today. Radar systems operate by emitting electromagnetic waves through a transmitter into the surrounding environment. These waves bounce off surfaces and return to a receiver, usually an antenna, which often serves both transmitting and receiving functions. For each wave, the receiver measures signals with different time delays, which are used to calculate distances to various objects [2]. One of radar's key advantages is its resilience in harsh weather conditions and its ability to function effectively in both day and night [14]. Since its inception, radar has undergone continuous refinement, becoming increasingly reliable. However, it still faces limitations. A common issue is the occurrence of ghost targets, false reflections caused by electromagnetic wave bounces that mimic real objects [15]. While such anomalies can be detected and filtered out using neural networks, radar also struggles in dense and complex environments due to its relatively low resolution. Some single-chip radar systems offer resolutions up to 100 times lower than lidar, making them less effective in cluttered scenes [16]. These limitations have led companies like Tesla to prioritize camera-based systems over radar. However, cameras are also prone to performance degradation under adverse conditions, such as fog, rain, or poor lighting. These constraints have hindered many current autonomous vehicles from reaching higher SAE levels of automation. In contrast, Waymo uses an integrated sensor suite, combining radar, lidar, and cameras, which has allowed its systems to achieve Level 4 automation [17]. This demonstrates the value of radar when paired with complementary technologies, particularly in overcoming environmental challenges and enhancing overall perception capabilities.

Nevertheless, this integration comes at a high cost. The estimated manufacturing cost of a fully equipped Waymo vehicle, including advanced lidar, radar, and cameras, can reach up to $300,000 [18]. Currently, the service is not profitable [19], but given that these systems are still in active development, there remains substantial potential for future optimization and scalability.

4. Lidar methods

Lidar technology uses light pulses, rather than electromagnetic waves, to map and interpret the surrounding environment. The concept of using light pulses for atmospheric measurement dates back to 1938, but the term *"lidar"* was not introduced until 1953 [3]. The invention of the laser in 1960 significantly accelerated lidar's development. Today, lidar is already proficient in meeting the perception needs of autonomous driving and continues to evolve at a rapid pace. Lidar systems operate by emitting laser light waves, which scatter and reflect upon encountering surfaces in the environment. As these light waves return to the sensor, they are detected by a photodetector, which converts the returning light into electrical signals for further processing. Using the Time-of-Flight principle, lidar measures depth based on the time it takes for light to travel to an object and back, similar to how radar measures distance using electromagnetic waves. After further calculations, the system generates 3D point cloud data to map the environment with high precision. Despite its advantages, lidar has several limitations. One common issue is reflection-based distortion, which can result in missing or corrupted data, perceived as “black holes” in the point cloud. Another challenge is its high cost and maintenance, which limits its practicality for large-scale deployment. While advanced algorithms and more affordable sensors are helping address these issues, the most critical drawback is lidar’s reduced performance in adverse weather conditions. Light scattering and occlusion due to fog, rain, or snow can degrade its accuracy [20].

Although lidar excels at generating detailed 3D environmental models under ideal conditions, expanding autonomous driving to a broader set of scenarios, especially weather-compromised ones, requires overcoming these limitations. Sensor fusion, particularly integrating lidar with radar, offers a promising solution. This approach combines the high spatial resolution of lidar with the robustness of radar in poor visibility, paving the way for autonomous vehicles to function reliably across a wider range of conditions and move closer to achieving SAE Level 5 autonomy.

5. Contrasting radar and lidar

Table 1 A summary of key differences between lidar and radar

	RADAR	**LIDAR**
Core Principle	Uses electromagnetic waves from a transmitter; measures reflected waves with a receiver to determine distances via signal delays.	Emits light pulses via a laser; uses Time-of-Flight principle to calculate distance and generate 3D point clouds.
Key Challenges	Ghost targets from false reflections. Low resolution in complex, dense environments.	Reflections may cause “black holes” or missing/distorted data in the point cloud. Poor performance in adverse weather.
Current Maturity	Highly reliable and widely used, millimeter-wave radar is improving resolution.	Rapid advancements in algorithms, sensor technology, and cost-reduction are accelerating adoption.

Radar and lidar are highly complementary sensing technologies in autonomous driving. While lidar offers high-resolution spatial data, it struggles in adverse weather conditions such as fog, rain, and snow. Radar, on the other hand, performs reliably in such environments but produces lower-resolution data. Lidar excels at capturing fine detail for 3D environmental mapping, though typically at shorter ranges. In contrast, radar’s longer range makes it effective for early object detection, a key capability for collision avoidance systems, even if object detail is limited.

This interplay makes radar an effective redundancy layer for lidar, ensuring robust system performance when visibility is compromised. Meanwhile, lidar contributes the precision needed for fine-grained perception and localization. By integrating radar and lidar through deep learning, autonomous systems can combine their respective strengths and compensate for their weaknesses. This sensor fusion results in a more resilient and adaptive perception system, capable of navigating a wider range of environmental and operational conditions with higher reliability.

6. Deep learning integration

Multimodal learning is a subfield of modern deep learning focused on solving complex problems by integrating multiple types of data. Its central goal is to fuse inputs from different modalities, data sources derived from distinct physical or informational processes. Just as humans combine inputs from sight, taste, and other senses to understand the world around them, machines can benefit from combining visual, textual, or sensory data to improve performance on specific tasks. For example, integrating image and text data can enhance the effectiveness of image search systems. In the context of this paper, the modalities of interest are lidar and radar, and the deep learning task at hand is object detection for autonomous vehicles. Applying multimodal learning to fuse radar and lidar data addresses the unique challenges of each sensor while maximizing their complementary advantages. Other sensor-based approaches for object detection have also been explored. Tesla’s Tesla Vision, for instance, relies solely on camera input and deep learning to build spatial awareness [21]. However, depending on a single sensor introduces a single point of failure, which increases the risk of error and decreases system reliability under variable conditions. By contrast, multimodal learning systems incorporating multiple sensors offer redundancy and fault tolerance, which are crucial for safety-critical applications like self-driving cars, where human lives are at stake. One notable approach for radar-lidar fusion is the Multi-Modal and Multi-Scale Fusion (M2-Fusion) framework [22]. This method includes two core components: the Center-based Multi-Scale Fusion (CMSF) module, which processes data at varying spatial resolutions, and the Interaction-based Multi-Modal Fusion (IMMF) module, which enhances feature-level interactions between the modalities. Together, these modules improve the system's ability to detect and represent diverse objects, ranging from small traffic cones and curbs to other vehicles, under real-world driving conditions.

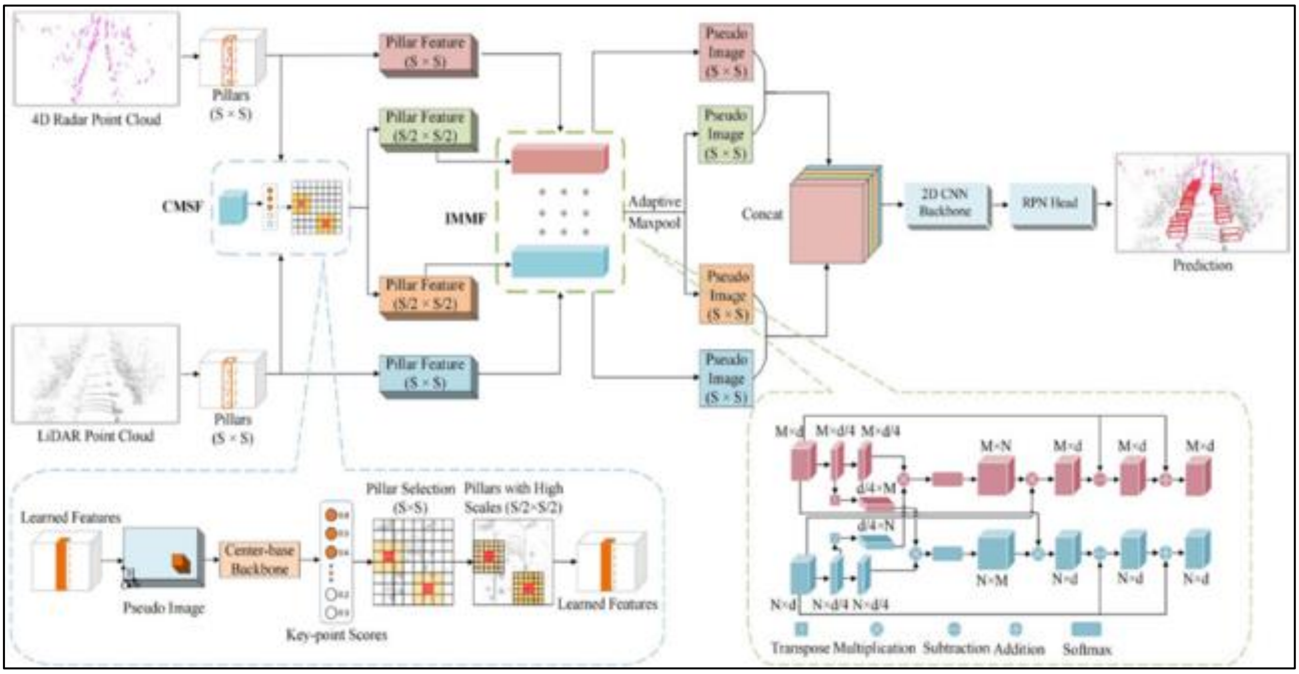

Figure 4 The structure of M2-Fusion contains the CMSF and the IMMF acting as the two main components[22]

The first module of M2-Fusion, known as the Center-based Multi-Scale Fusion (CMSF), is designed to enhance object detection, particularly in the presence of sparse radar data. Because radar often lacks the resolution and point density of lidar, its raw output is insufficient for high-precision tasks required in autonomous driving. To address this, CMSF begins by identifying and extracting key features from the available radar data. A voxel-based framework is then applied to structure the data. Voxels, essentially the 3D equivalent of 2D pixels, divide the point cloud into a grid structure for spatial organization (see Figure 5). However, a major challenge lies in selecting the appropriate voxel grid size. Larger voxels reduce computational load but sacrifice fine detail, while smaller voxels preserve detail at the cost of increased noise and processing time.

To strike a balance, CMSF employs multi-scale voxelization, using a range of voxel sizes to capture both coarse and fine-grained information. To minimize computational overhead, the module first identifies important anchor points in the radar data, then generates voxel grids selectively around those points. This approach ensures that the most salient features are extracted without exhausting system resources, thereby improving the overall quality and efficiency of object detection.

Figure 5 NVIDIA wins the 3D Occupancy Prediction Challenge with FB-OCC, a technology designed for using camera data to generate 3D voxel bird's-eye maps. The bottom two maps exemplify the creation of voxel grids using sensor data from the six camera inputs above [23]

Once key features are extracted, the Interaction-based Multi-Modal Fusion (IMMF) module integrates and aligns corresponding attributes across the lidar and radar modalities. These attributes may include an object's position, volume, and surface geometry. IMMF plays a critical role in establishing meaningful correlations between the data types, effectively bridging the gap between the strengths and weaknesses of each sensor. This fusion becomes especially valuable in adverse weather conditions. For example, in foggy environments, the lidar features extracted by CMSF may degrade in quality compared to those captured in clear conditions. The IMMF compensates by enriching the lidar-derived features using the more reliable radar data, which remains largely unaffected by fog. This fusion ensures that even under challenging circumstances, the object's key characteristics are preserved and accurately interpreted.

The performance and output of the M2-Fusion model, including the contributions of the IMMF, are illustrated in Figure 6.

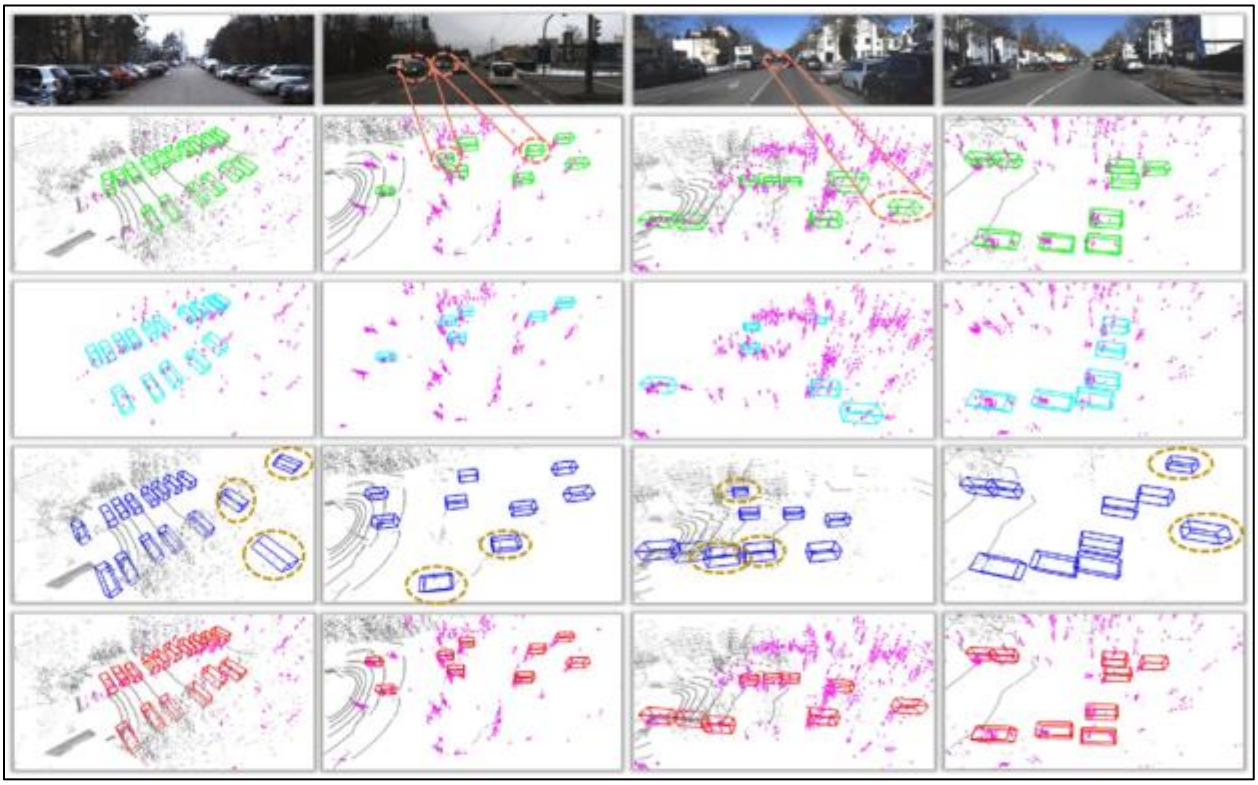

Figure 6 The qualitative results produced by the M2-Fusion model using the Astyx HiRes 2019 dataset. The first row displays RGB images of the driving scenes. The second row shows the ground truth annotations, where green bounding boxes highlighted by dotted orange ovals indicate missed detections. The third row illustrates object detection results using radar data alone, while the fourth row shows results using lidar only. In the lidar results, blue bounding boxes surrounded by dotted brown ovals represent false positives. Finally, the last row demonstrates the outputs of the M2-Fusion approach, where the fusion of radar and lidar data leads to more accurate detections by reducing both missed detections and false positives

In extreme scenarios where one of the two sensors, either radar or lidar, experiences a complete failure, the M2-Fusion model is unable to support fully autonomous navigation. Since the IMMF module depends on synthesizing data from both modalities, its functionality is compromised when one input source is lost, resulting in a significant decline in detection performance. M2-Fusion's architecture inherently assumes that both sensors are operational. However, real-world factors such as sensor aging, wear and tear, manufacturing defects, or physical damage can lead to unexpected sensor failures [24].

To address this vulnerability, alternative methods have been proposed to ensure autonomous driving systems can maintain performance even when one modality is compromised. As illustrated in Figure 7, one such approach is the Self-Training Multimodal Vehicle Detection Network (ST-MVDNet), which has been designed to support asymmetric sensor availability and adaptively continue operation despite missing inputs [25].

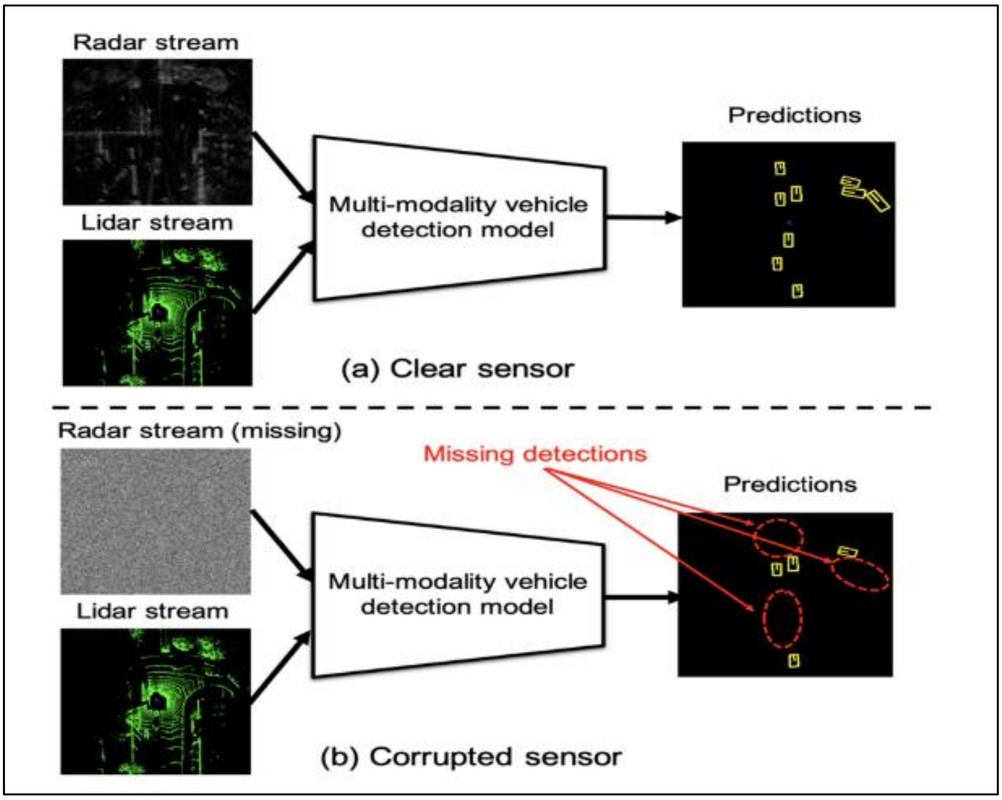

Figure 7 When one data stream unexpectedly becomes inaccessible, the multimodal detection model is unable to identify all objects. ST-MVDNet has been proposed to mitigate such an issue[25]

ST-MVDNet is built upon the foundation of the earlier Multimodal Vehicle Detection Network (MVDNet) [26] and is trained using the Mean Teacher framework [27]. Like M2-Fusion, the original MVDNet was developed to integrate lidar and radar data, leveraging the complementary nature of the two modalities under the assumption that both sensors are always available. MVDNet employs a two-stage architecture to extract and fuse important features. In the first stage, MVDNet identifies object proposals, regions likely to contain vehicles, from both radar and lidar data independently. In the second stage, these proposals are fused using 3D convolutional layers and attention mechanisms that focus on the most informative regions across the combined feature space. This selective attention improves detection accuracy by emphasizing meaningful patterns while filtering out noise. The performance and effectiveness of MVDNet are illustrated in Figure 8. However, it is important to note that MVDNet still requires both modalities for optimal results, and like M2-Fusion, it is not robust to scenarios where one sensor fails or is missing.

ST-MVDNet applies the Mean Teacher framework by utilizing a teacher–student model architecture, as illustrated in Figure 9. In this setup, the teacher model generates predictions on a dataset, which are then used to compute a consistency loss, a measure of how consistent the predictions remain when the same input is subjected to different conditions or perturbations. This consistency loss guides the student model during training, prompting it to update its parameters to better align its predictions with those of the teacher.

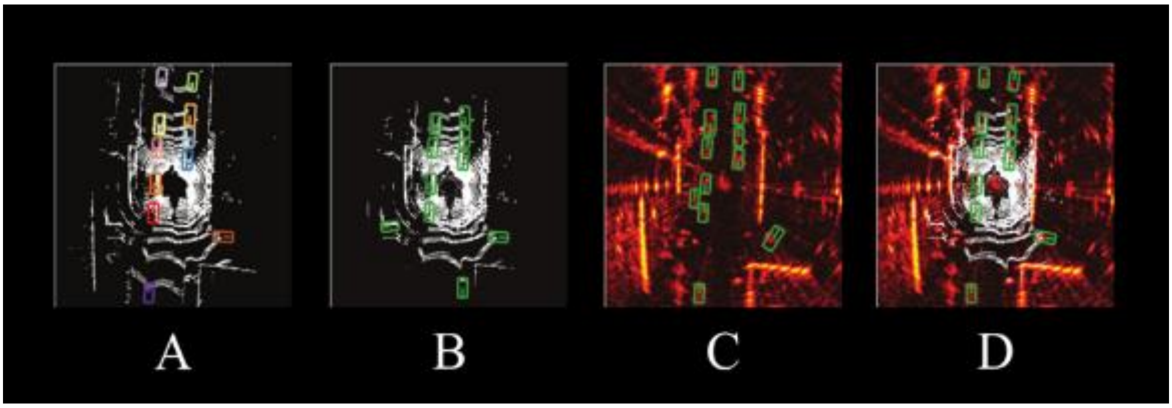

Figure 8 The object detection results of the MVDNet framework. Figure A provides a top-down view of the 3D lidar point cloud, serving as the ground truth, with each colored bounding box representing a detected vehicle. The central vehicle is equipped with both lidar and radar sensors. Figure B shows MVDNet's performance using lidar-only data, where vehicles at the farthest range are missed due to fog interference, and some background points are misclassified as vehicles. Figure C presents results using radar-only data, which leads to false positives caused by radar noise. In contrast, Figure D demonstrates the effectiveness of fusing lidar and radar data through deep learning in MVDNet, successfully detecting vehicles with high accuracy, closely matching the ground truth provided in Figure A [26]

The interaction between the teacher and student models is central to the success of ST-MVDNet. While the teacher model is provided with complete and clear input modalities (i.e., both radar and lidar), the student model is intentionally trained with missing or corrupted sensor inputs, such as absent radar or lidar streams. This design enables the student network to develop robustness to sensor failure and enhances generalization by avoiding overfitting, a condition in which the model becomes too closely tailored to the training data and fails to perform well on unseen inputs.

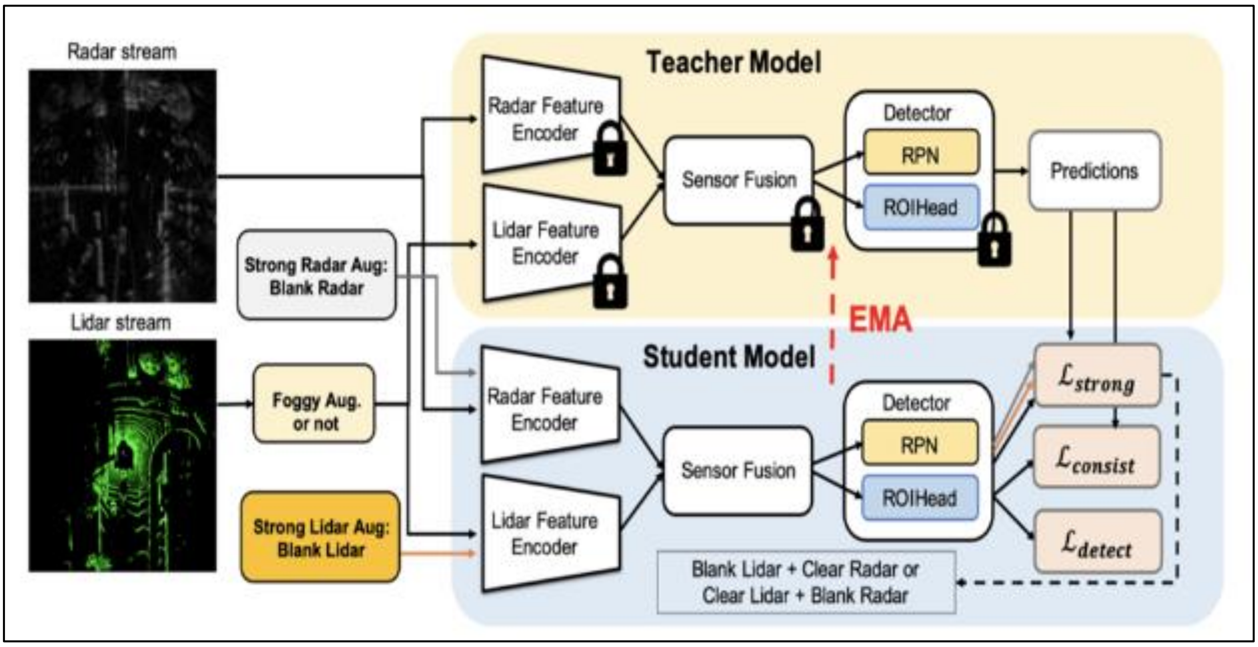

Figure 9 The framework of ST-MVDNet, built upon the Mean Teacher principle and employs a teacher student model pair. Both models share the same architecture and include two feature encoders that convert raw lidar and radar data into a format suitable for further processing. Each model also contains a sensor fusion module to integrate features across modalities, as well as a detector that generates feature maps and produces object proposals for vehicle detection [25]

The results of ST-MVDNet are presented in Figure 10, alongside a comparative analysis with MVDNet. When trained without strong data augmentation, ST-MVDNet tends to produce false positives and fails to detect all objects within the scene. This highlights a potential drawback: the model's dependency on augmented data for optimal performance. Such augmentation, while beneficial during training, can differ significantly from real-world conditions and may lead to overfitting or increase the computational complexity of training.

However, when robust augmentation techniques are incorporated, ST-MVDNet demonstrates a marked improvement in object detection accuracy. Remarkably, the model can correctly identify objects even in the absence of one data stream (either radar or lidar), underscoring its resilience and fault tolerance capabilities that are essential for reliable autonomous vehicle operation.

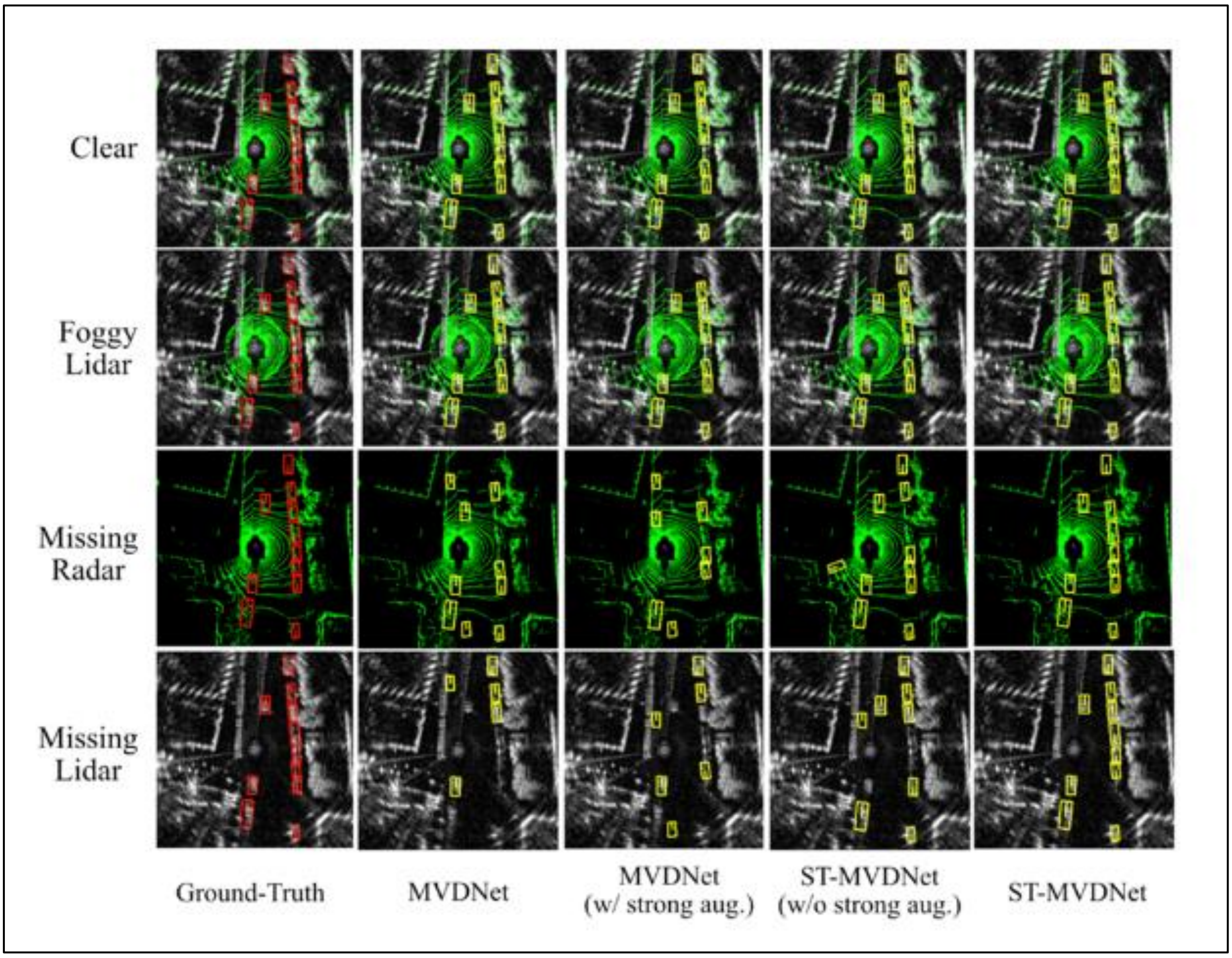

Figure 10 The detection results using yellow bounding boxes to indicate objects identified in the scene. During testing with missing radar or missing lidar data streams, both MVDNet and ST-MVDNet (when trained without strong augmentation) exhibit missed detections and false positives. However, when strong data augmentation is applied during training, ST-MVDNet demonstrates robust performance, successfully detecting objects even in the absence of one sensor input, and producing accurate and consistent results across all tested scenarios [25]

7. Conclusion

The primary aim of this paper was to explore various methods of integrating lidar and radar for object detection in autonomous vehicles. We began by examining the historical development and current capabilities of radar, lidar, and deep learning, establishing a foundation for understanding their roles in modern computer vision systems. Given the complementary strengths of radar and lidar, we highlighted the power of sensor fusion as a means to enhance perception in autonomous driving. To investigate this further, we analyzed two prominent multimodal learning frameworks, M2-Fusion and ST-MVDNet, focusing on their methodologies, performance outcomes, and limitations. Such critical evaluation is essential when considering whether to adopt new technologies into real-world manufacturing systems. Our findings suggest that while both models demonstrate promising object detection capabilities, they remain constrained by sensor dependency and limited training generalization. In scenarios where sensor input is missing or when training lacks sufficient augmentation, both systems fail to maintain the reliability required for guaranteeing passenger safety.

Nevertheless, if perfected, multimodal learning using lidar and radar holds immense potential to impact a broad range of industries and improve safety in autonomous vehicles. Current research provides a solid foundation, but further work is needed to broaden the applicability of these models. Future research should explore strategies such as larger and more diverse datasets, model regularization, and k-fold cross-validation to overcome the limitations observed in current implementations.

In conclusion, although current radar-lidar fusion systems require further development for robust real-world deployment, their demonstrated capabilities mark a significant step toward advancing autonomous driving technologies. The research reviewed here contributes to this progress, bringing us closer to a future of safer, more intelligent road transportation.

References

[1] IBM. What is deep learning? 2023. Available from: https://www.ibm.com/topics/deep-learning

[2] Srivastav A, Mandal S. Radars for Autonomous Driving: A Review of Deep Learning Methods and Challenges. arXiv:2306.09304 [cs.CV]; 2023.

[3] Wandinger U. Introduction to lidar. In: Lidar: Range-resolved optical remote sensing of the atmosphere. Springer; 2005. p. 1–18.

[4] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5:115–133.

[5] Chandra AL. McCulloch-Pitts Neuron, Mankind's First Mathematical Model Of A Biological Neuron. 2018. Available from: https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1

[6] IBM. What is a neural network? 2023. Available from: https://www.ibm.com/topics/neural-networks

[7] SAE International. SAE Levels of Driving AutomationTM Refined for Clarity and International Audience. 2021. Available from: https://www.sae.org/blog/sae-j3016-update

[8] Tesla. AI & Robotics. 2023. Available from: https://www.tesla.com/AI

[9] Chougule A, et al. A Comprehensive Review on Limitations of Autonomous Driving and Its Impact on Accidents and Collisions. IEEE Open J Veh Technol. 2024;5:142–161. doi:10.1109/OJVT.2023.3335180

[10] Cristovao N. All Tesla FSD Visualizations and What They Mean. 2022. Available from: https://www.notateslaapp.com/tesla-reference/636/all-tesla-fsd-visualizations-and-what-they-mean

[11] Muhammad K, et al. Deep Learning for Safe Autonomous Driving: Current Challenges and Future Directions. IEEE Trans Intell Transp Syst. 2021;22(7):4316–4336. doi:10.1109/TITS.2020.3032227

[12] Dremio. Adversarial Attacks in AI. 2024. Available from: https://www.dremio.com/wiki/adversarial-attacks-in-ai/

[13] Lyft. Dash Lyft Perception. 2024. Available from: https://dash.gallery/dash-lyft-explorer/

[14] Waymo. Waymo Safety Report. 2020. Available from: https://storage.googleapis.com/sdc-prod/v1/safety-report/2020-09-waymo-safety-report.pdf

[15] Jeong T, Lee S. Ghost Target Suppression Using Deep Neural Network in Radar-Based Indoor Environment Mapping. IEEE Sens J. 2022;22(14):14378–14386. doi:10.1109/JSEN.2022.3182377

[16] Prabhakara A, et al. High Resolution Point Clouds from mmWave Radar. In: IEEE Int Conf Robot Autom (ICRA); 2023. doi:10.1109/icra48891.2023.10161429

[17] Webb N, et al. Waymo's Safety Methodologies and Safety Readiness Determinations. arXiv:2011.00054 [cs.RO]; 2020.

[18] LeVine S. What it costs to turn a car into a self-driving vehicle. 2017. Available from: https://qz.com/924212/what-it-really-costs-to-turn-a-car-into-a-self-driving-vehicle

[19] Broadwater B. Waymo Stock – Is Waymo Publicly Traded in 2024? 2017. Available from: https://www.wealthdaily.com/waymo-stock/

[20] Dreissig M, et al. Survey on LiDAR Perception in Adverse Weather Conditions. arXiv:2304.06312 [cs.RO]; 2023.

[21] Tesla. Tesla Vision Update: Replacing Ultrasonic Sensors with Tesla Vision. 2021. Available from: https://www.tesla.com/support/transitioning-tesla-vision

[22] Wang L, et al. Multi-Modal and Multi-Scale Fusion 3D Object Detection of 4D Radar and LiDAR for Autonomous Driving. IEEE Trans Veh Technol. 2023;72(5):5628–5641. doi:10.1109/TVT.2022.3230265

[23] Li Z, et al. FB-OCC: 3D Occupancy Prediction based on Forward-Backward View Transformation. arXiv:2307.01492 [cs.CV]; 2023.

[24] Safavi S, et al. Multi-Sensor Fault Detection, Identification, Isolation, and Health Forecasting for Autonomous Vehicles. 2021. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8038547/

[25] Li YJ, et al. Modality-agnostic learning for radar-lidar fusion in vehicle detection. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit. 2022. p. 918–927.

[26] Qian K, et al. Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit. 2021. p. 444–453.

[27] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Adv Neural Inf Process Syst. 2017;30.